A DISCIPLINE FOR THE AVOIDANCE OF UNNECESSARY ASSUMPTIONS

LEWIS H. ROBERTS

VOLUME LIV, PAGE 205

DISCUSSION BY ROBERT L. HURLEY

Lewis H. Roberts' paper, "A Discipline for the Avoidance of Unnecessary Assumptions," published in the CAS *Proceedings* Volume LIV (1967) was initially prepared for the seminar presented by the Committee on Mathematical Theory of Risk at the Society's meetings in Detroit, Michigan, November 1966. While the presentations of this colloquium have been made available to the membership in a separate booklet, it is a happy stroke that the Roberts' paper will appear in the official *Proceedings* as readily accessible research material for present and future students of the insurance business.

In his introductory section, Roberts was quick to eschew any thought of venturing into a philosophical treatise on "Ockham's razor." This reference was offered solely as a citation of some pertinency to color a fairly recent mathematical development of likely promise to the actuarial profession. In essence, the paper advances the idea that E. T. Jaynes' Formalism, developed from the Shannon treatment of "Entropy" or "Uncertainty," may possibly be applied to certain situations not uncommonly encountered in actuarial work.

For example, the actuary on occasions has only various averages culled from the data rather than complete information on the frequency distributions of the losses and/or premiums and/or exposures to work with. Naturally, there is some concern that we do not read into such available information more than we truly have. And in paraphrase of Ockham's admonition, "Don't search for a more elaborate rationalization to account for a particular phenomenon than you need to explain the basic facts concisely." As Roberts indicated in comparing various statements about a distribution, the principle of maximum entropy helps us to select the one involving the minimum subjectivity on the part of the observer. The larger the number of alternatives available to interpret some observation, the less sure the observer can be that he has chosen the most appropriate explanation. Truly, there are inescapable polarities between our knowledge and our ignorance. The certainty to be accorded to our information seems to vary inversely with the precision to be attached thereto.

It appears to this reviewer that Roberts has forcefully identified the source and clearly traced the consequences of bias and prejudiced data.

The CAS Detroit seminar was undoubtedly arranged for the guidance of the general membership who, like this reviewer, may not be particularly conversant with recent mathematical developments. And it may have been a singularly felicitous adventure that Roberts chose to further the research being done on information theory by scientists with prime allegiance to various disciplines other than actuarial work.

The current activity on the information theory is believed to stem from Dr. Claude E. Shannon's 1948 paper in the *Bell System Technical Journal* which was concerned with developing a statistical theory of the information sum from successive units originating as individual decisions from equally probable choices. At about the same time, R. A. Fisher was investigating a similar idea from the view of classical theory of statistics, and Norbert Wiener was founding the field of Cybernetics from a parallel source. In his reminiscences Dr. Wiener relates that although Claude Shannon was a student during his teaching days at MIT and that they then, and later, had occasions to discuss scientific matters, their respective work in information theory, as far as he could recall, was developed independently.

The Shannon initial monograph, "A Mathematical Theory of Communication," was later supplemented with an essay by Warren Weaver, and published in book form by the University of Illinois Press 1949. It is still readily available from most large libraries. The Shannon contribution does not make easy reading for the uninitiate. This reviewer struggled, with more exasperation than success, over Appendix 2 which gives the mathematical derivation of the Shannon Equation $H = -\sum p_i \log p_i$. Not that the mathematics would be impossibly difficult for the average competence required for most actuarial research. Yet it might conceivably have appeared to some that Shannon had sharpened his intuitive skills in mathematics so as to suggest a reasonable degree of contempt for those who may prefer not to flash from one intellectual peak to the next.

With the above experience, it was somewhat heartening to chance upon the observation in A. I. Khinchin's *Mathematical Foundations of Information Theory,* Dover 1962, that while Shannon was a highly competent scientist whose discoveries in information theory were truly remarkable, the mathematical display of the findings seemed to lack at certain points the rigor and clarity that one could wish for. Unfortunately, in his own development, Khinchin, while engaging in a somewhat more detailed refinement of the mathematics, seemed to parallel the general tenor of the Shannon exposition without a compensating gain in the lucidity of the mathematical argument. It is understandable, therefore, that Roberts would be satisfied just to identify Shannon's equation, and the basic criteria on which it was founded, and to offer certain observations thereon without detailing either the Shannon or Khinchin mathematical involvements — or as they may appear to any of us who might qualify as less sophisticated readers.

Roberts, it seems to this reviewer, properly highlighted the applications that have been made from the original information theory in the field of thermodynamics and thermostatics. He credits Myron Tribus with various contributions to the gradual realization of the possible extensions of the underlying concepts, and cites a number of Tribus' articles and technical papers thereon. This reviewer would like to add one further reference to Roberts' list, Tribus' text book *Thermostatics and Thermodynamics,* published by D. Van Nostrand 1961, wherein ingenuously simple and delightfully straightforward derivations of Shannon's equations are given. These are considered prerequisites to an understanding of the basic principles of information theory and its possible application to actuarial problems. Consequently they have been sketched out in an appendix to this review.

In his paper Roberts notes that often the only information available to the actuary is the average readings for some variable whose underlying loss distribution is unknown. He suggests that an extension of E. T. Jaynes' Formalism may enable the investigator to select that one distribution which affords the maximum entropy according to the Shannon development. The equation for the average reading is given in Roberts' paper as

$\overline{g}_r(x) = \sum p_i g_r(x_i)$, where $r = 1, 2, 3 \ldots m$ for m functions and $\sum p_i = 1$.

Roberts noted that the assignment of the p_i for which S is a maximum is given by the equation:

$$p_i = exp.\ [-a_o - a_1 g_1(x_i) - a_2 g_2(x_i) - \ldots.]$$

with

$$a_o = ln \sum_i exp.\ [\sum_r a_r g_r(x_i)]$$

$$\overline{g}_r = -\,\delta a_o / \delta a_r$$

where the "a"s are the Lagrangian multipliers satisfying the requirements of $\overline{g}_r(x)$.

On pages 70 through 74 in his book *Thermostatics and Thermodynamics* Tribus gives a good explanation of the Lagrangian multipliers and examples of their use. Somewhat later in the same chapter he outlines a fairly complete derivation of the equations underlying Jaynes' Formalism, which is one starting point of Roberts' further developments. While the mathematics are too extensive to attempt to sketch out in his review, it is believed that they are not beyond the competence required of CAS members. The interested actuary will undoubtedly find that some extra effort thereon will prove worthwhile.

It is this reviewer's belief that an author in any scientific inquiry has discharged his obligation to his readers and to his own conscience if he has advanced a logically consistent proposal or theory, examined its ramifications in the conceptual framework presently conditioning the particular field, and suggested aspects or areas in which future research may prove promising. He is not in conscience required to present a bill of particulars on the many associated details for some future implementation, although he may do so, obviously, if he chooses.

In citing three specific problems where the Roberts-Jaynes extension of the Shannon entropy concept might be used, Roberts successfully blended the daring often founded in the creative thinker with the caution associated with the successful business man. For example, while offering these equations as a method for computing deductible credits when, say, only the average loss is known, he prudently questions the assignments that would likely be made in selecting the value for the coefficients of the exponents — and inquires whether any such selection might not itself betray a prejudice.

In the particular area of loss distributions, insurance research often has more than average loss size to work with. In certain individual studies the problem has seemed to be not primarily a matter of the degree of the detail. Rather, our main problem has been, in such instances, to develop an adequate mathematical relationship among the variables so that we may interpolate readings for which direct computations are not provided by the statistics. In any such situation, the author notes that if we have more information, we should use it, since the equations in his paper apply when the data are available only in the form of expected values.

In the area of property insurance, it would appear a somewhat hazardous venture to posit a frequency distribution of losses and amounts of insurance by size in order to develop Loss Elimination Ratios for a deductible rating plan — knowing only the average value and/or the range of values. Admittedly, the underlying equation might be expected to be of the exponential form with negative exponents. But such fragmentary knowledge, while possibly of some value in the absence of other data, may be not much, if any, advantage over the intuitive skills of some knowledgeable underwriter exercising his judgment as to what a given deductible would be worth, rate wise, on a specific class of business.

Roberts also notes in his paper that this "entropy" concept might also be used in the planning of risk classification plans for rate differential purposes and in evaluations of credibility with regard to the probability distribution of error in the existing rate levels. He observes that the classification plan with the smallest entropy value would afford the most information; and conversely, the entropy value would be greatest for the most homogeneous population. He then cites the work done by R. A. Bailey (*PCAS* Volume XLVII — 1960) on classification analysis using the coefficient of variation technique and concludes that both approaches would likely afford answers of about the same order of magnitude.

It was interesting that the author would speculate that, unlike the earlier statistical techniques, the method of maximum entropy would not provide for any comparison of the hypothesis with observed events. He advises that no such testing is possible since the method uses all pertinent information available. He cites the parallel with Bayes theorem wherein solutions are complete and final and allow of no further referrals. He contrasted it with the Neyman-Pearson tradition of testing hypotheses and delimiting regions within which the "true" answer might be expected most probably to lie.

It would be understandable that the reader might entertain some misgivings on an approach somewhat strange in the light of his previous experience in the testing of hypotheses. It is possible that his uneasiness might be due only in part to the consideration that the technique may represent a break with statistical tradition. There are men of some stature in statistical theory who regard the current Bayesian trend, if not a break, at least as estrangement with statistical realities. Maybe in view of the responsibilities with which the actuary is charged, he must necessarily examine any novel proposals with a fair degree of circumspection. At the same time, he cannot afford to be indifferent to new ideas and neglect the developments that are taking place in associate disciplines. This reviewer believes, therefore, that the Roberts paper represents a valuable addition to our *Proceedings*.

APPENDIX TO DISCUSSION BY HURLEY

In terms of the criteria in Roberts' paper, the entropy or uncertainty (S):

(a) should depend only on probability distribution;

$$\therefore S = f(p_1, p_2, \ldots . p_n)$$

(b) should be monotonic function of "n," if $p_1 = p_2 = p_3 \ldots . = p_i$

(c) if W and Y are independent events and Z is a compound event of W and Y, then the uncertainty about Z should be defined as the sum of the separate events' uncertainties, or if

$$W \cdot Y = Z, \text{ then } S(Z) = S(W) + S(Y).$$

If the numerical value of "S" must be independent of the way the problem is set up, then criterion *(b)* requires $S = f(n)$, when $p_i = \frac{1}{n}$ for each "i," and criterion *(c)* requires $f(x^m) = mf(x)$ and Shannon proved that the only function satisfying this relationship is:

$f(x) = k \ln x$ where k is constant; the Tribus proof is:

(1) $f(x^m) = mf(x)$, differentiating with respect to m letting $(x^m) = U$

Left Side $= \frac{df}{du} \cdot \frac{du}{dm} = \frac{df}{du} \cdot x^m \ln x \frac{dm}{dm}$; Right Side $= f(x) \frac{dm}{dm}$;

or (2) $\frac{df}{du} \cdot x^m \ln x = f(x)$, and now differentiating (1) by x,

Left Side $= \frac{df}{du} \cdot \frac{du}{dx} = \frac{df}{du} \cdot mx^{m-1} \frac{dx}{dx}$; Right Side $= mf'(x)\frac{dx}{dx}$;

or (3) $\frac{df}{du} \cdot mx^{m-1} = mf'(x)$ where $f'(x) = \frac{df(x)}{dx}$.

Next eliminate df/du from equations (2) and (3), and

(4) $\frac{f(x)}{x \ln x} = f'(x)$, or

(5, a) $\frac{f'(x)}{f(x)} = \frac{1}{x \ln x}$; or (5, b) $\frac{df(x)}{f(x)} = \frac{dx}{x \ln x}$ $\left(\text{since } f'(x) = \frac{df(x)}{dx}\right)$;

(6) $\int \frac{df(x)}{f(x)} = ln\ f(x) + C_1$ and

(7) $\int \frac{dx}{x\ ln\ x} = ln(ln\ x) + C_2$; therefore

(8, a) $ln\ f(x) + C_1 = ln(ln\ x) + C_2$; or (8, b) $f(x) = k\ ln\ x$

From criterion (b) above, $S = f(n)$ is a monotonically increasing function of n. When all the "p_i"s are equal, equation (8, b) gives $S = k\ ln(n)$ with equal "p_i"s, each $p_i = \frac{1}{n}$ and

(9) $S = k\ ln\left(\frac{1}{p_i}\right)$ or

(10) $S = -\ k\ ln\ p_i$